



Posets and Permutations in the duplication-loss model

Mathilde Bouvel, Elisa Pergola

► To cite this version:

Mathilde Bouvel, Elisa Pergola. Posets and Permutations in the duplication-loss model. GAS-COM'08, Jun 2008, Bibbiena, Italy. <hal-00322737>

HAL Id: hal-00322737

<https://hal.archives-ouvertes.fr/hal-00322737>

Submitted on 18 Sep 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Posets and Permutations in the duplication-loss model

Mathilde Bouvel ^{*} Elisa Pergola [†]

Abstract

In this paper, we are interested in the combinatorial analysis of the whole genome duplication - random loss model of genome rearrangement initiated in [8] and [7]. In this model, genomes composed of n genes are modeled by permutations of the set of integers $[1..n]$, that can evolve through duplication-loss steps. It was previously shown that the class of permutations obtained in this model after a given number p of steps is a class of pattern-avoiding permutations of finite basis. The excluded patterns were implicitly described as the minimal permutations with $d = 2^p$ descents, minimal being intended in the sense of the pattern-involvement relation on permutations. Here, we give a local and simpler characterization of the set \mathcal{B}_d of minimal permutations with d descents. We also provide a more detailed analysis - characterization, bijection and enumeration - of a particular subset of \mathcal{B}_d , namely the patterns in \mathcal{B}_d of size $2d$.

1 Pattern-avoidance in the duplication-loss model

The study of genome evolution has been the source of extensive research in computational biology in the last decades. Many models for genome evolution were defined, taking into account various biological phenomena (see [4], [9], [10] for recent examples in the literature). Among them, the *tandem duplication - random loss model* represents genomes with permutations, that can evolve through *duplication-loss steps* representing the biological phenomenon that duplicates fragments of genomes, and then loses one copy of every duplicated gene. For the original biological motivations, we refer to [8]. In this first section, we describe the duplication-loss model, and recall some previous results obtained by other authors. We recall some definitions and properties on pattern-avoidance that are necessary to introduce the permutations that will arise from this model and on which we will focus in the rest of the paper.

^{*}LIAFA, Université Paris Diderot - Paris 7, Case 7014, 75205 Paris Cedex 13, France, mbouvel@liafa.jussieu.fr, supported by ANR project number 07-2.195422 (GAMMA)

[†]Dipartimento di Sistemi e Informatica, v.le Morgagni 65, 50134 Firenze, Italy, elisa@dsi.unifi.it

1.1 The whole genome duplication - random loss model

A permutation of size n is a bijective map from $[1..n]$ to itself. We denote by S_n the set of permutations of size n . We consider a permutation $\sigma \in S_n$ as the word $\sigma_1\sigma_2\ldots\sigma_n$ of n letters on the alphabet $\{1, 2, \ldots, n\}$, containing exactly once each letter (we often prefer the word *element* instead of letter). For example, 346251 represents the permutation $\sigma \in S_6$ such that $\sigma_1 = 3, \sigma_2 = 4, \ldots, \sigma_6 = 1$.

The *whole genome duplication - random loss model* we take into consideration is a particular case of the more general *tandem duplication - random loss model*, with various cost functions, studied in [8] and [7]. It is worth noticing that these models of evolution of permutations can be viewed in the more general framework of *permuting machines* defined in [1], where properties concerning pattern-avoidance are already obtained.

In our model, permutations can be modified by *duplication-loss steps*. Each of these steps is composed of two elementary operations. Firstly, the permutation (a fragment of consecutive elements of it for the tandem duplication) is duplicated, and the duplicated copy is inserted immediately after the original copy: this is the *whole genome duplication*. Then the *random loss* occurs: one copy of every duplicated element is lost, so that we get a permutation at the end of the step.

1 2 3 4 5 6 \rightsquigarrow 1 2 3 4 5 6 1 2 3 4 5 6 \rightsquigarrow ~~1~~2~~3~~4 5~~6~~1~~2~~3~~4~~5 6 \rightsquigarrow 2 4 5 1 3 6

Figure 1: Example of one step of whole genome duplication - random loss

We focus on permutations obtained from an identity permutation $12\ldots n$ after at most a certain number p of duplication-loss steps in the whole genome duplication - random loss model. We will describe combinatorial properties of those permutations in Subsection 1.3, in terms of pattern-avoidance.

1.2 Previous result on the duplication-loss model

The permutations obtainable in at most p duplication-loss steps in the whole genome duplication - random loss model were implicitly characterized in [8]. In [7], using the classical definition of *descents* in permutations, we reformulated their result into Theorem 1.

Definition 1. *Given a permutation σ of size n , we say that there is a descent (resp. ascent) at position i , $1 \leq i \leq n - 1$, if $\sigma_i > \sigma_{i+1}$ (resp. $\sigma_i < \sigma_{i+1}$). We indicate the number of descents of the permutation σ by $\text{desc}(\sigma)$.*

Example 1. For example, $\sigma = 698413725$ has 4 descents, namely at positions 2, 3, 4, 7.

Theorem 1. The permutations that can be obtained in at most p steps in the whole genome duplication - random loss model are exactly those whose number of descents is at most $2^p - 1$.

Generalizing a little, we will focus in the remaining of the paper on the set of permutations with at most d descents, without assuming that d is of the form $d = 2^p - 1$. We can notice that this corresponds to the set of permutations composed of $d + 1$ increasing sequences, separated either by ascents or by descents (a permutation may have more than one such decomposition). In [2] this set is denoted $W(e_1, \dots, e_{d+1})$ with $\forall i, e_i = +$. In this paper, and as an application of their results, the authors are concerned with properties of $W(e_1, \dots, e_{d+1})$ in terms of pattern-avoidance, and they prove that this set is a finitely based pattern-avoiding permutation class. Our work can be seen as a more detailed analysis of this particular result.

In order to have a better view of descents and ascents in permutations, we consider their *grid representation* defined in [5] and described in Figure 2.

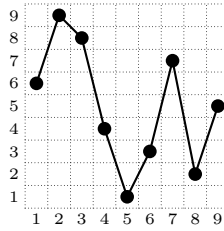


Figure 2: The grid representation of the permutation $\sigma = 698413725$

1.3 Pattern-avoidance in the duplication-loss model

Definition 2. A permutation $\pi \in S_k$ is a pattern of a permutation $\sigma \in S_n$ if there is a subsequence of σ which is order-isomorphic to π ; i.e. if there is a subsequence $\sigma_{i_1}\sigma_{i_2}\dots\sigma_{i_k}$ of σ (with $1 \leq i_1 < i_2 < \dots < i_k \leq n$) such that $\sigma_{i_\ell} < \sigma_{i_m}$ whenever $\pi_\ell < \pi_m$.

We also say that π is involved in σ and call $\sigma_{i_1}\sigma_{i_2}\dots\sigma_{i_k}$ an occurrence of π in σ .

Example 2. For example $\sigma = 142563$ contains the pattern $\pi = 1342$; and 1563, 1463, 2563 and 1453 are the occurrences of this pattern in σ . But σ does not contain the pattern 321 as no subsequence of size 3 of σ is isomorphic to 321, i.e., is decreasing.

We write $\pi \prec \sigma$ to denote that π is a pattern of σ . We say that a set \mathcal{C} of permutations is *stable for \prec* if, for any $\sigma \in \mathcal{C}$, for any $\pi \prec \sigma$, then we also have $\pi \in \mathcal{C}$.

A permutation σ that does not contain π as a pattern is said to *avoid* π . The class of all permutations avoiding the patterns $\pi_1, \pi_2, \dots, \pi_k$ is denoted $S(\pi_1, \pi_2, \dots, \pi_k)$. We say that $S(\pi_1, \pi_2, \dots, \pi_k)$ is a class of pattern-avoiding permutations of *basis* $\{\pi_1, \pi_2, \dots, \pi_k\}$. The basis of a class of pattern-avoiding permutations may be finite or infinite. Pattern-avoiding permutation classes considered in the literature (see for example [6], [12] and their references) are often of finite basis.

Although it may sound a powerful statement, it is simple to understand that:

Proposition 1. *A set \mathcal{C} of permutations that is stable for \prec is a class of pattern-avoiding permutations. However, its basis $\mathcal{B} = \{\sigma \notin \mathcal{C} : \forall \pi \prec \sigma \text{ with } \pi \neq \sigma, \pi \in \mathcal{C}\}$ might be infinite.*

In [7], we proved that classes of permutations defined in duplication-loss models, as the permutations obtained in at most a given number p of steps, are classes of pattern-avoiding permutations. We are not always able to find the basis, even though we have proved that this basis is finite. In this paper, we take into consideration in particular the following result:

Theorem 2. *The class of permutations obtainable in at most p steps in the whole genome duplication - random loss model is a class of pattern-avoiding permutations whose basis \mathcal{B}_d is finite and is composed of the minimal permutations with $d = 2^p$ descents, minimal being intended in the sense of \prec .*

The proof of Theorem 2 we gave in [7] is implicate, and for sake of clarity we give below an explicit proof of it.

Proof. Let us denote by \mathcal{C}_p the class of permutations obtainable in at most p steps in the whole genome duplication - random loss model.

We first prove that \mathcal{C}_p is stable for \prec . Consider $\sigma \in \mathcal{C}_p$ of size n and π of size $k \leq n$ such that $\pi \prec \sigma$. In a duplication-loss scenario from $12 \dots n$ to σ – containing at most p steps –, if you keep track only of the elements that form an occurrence of π in σ , you obtain a sequence of duplication-loss steps moving from $12 \dots k$ to π , of no more than p steps. This shows that $\pi \in \mathcal{C}_p$, and consequently, that \mathcal{C}_p is stable for \prec .

According to Proposition 1, \mathcal{C}_p is a class of pattern-avoiding permutations whose basis is $\{\sigma \notin \mathcal{C}_p : \forall \pi \prec \sigma \text{ with } \pi \neq \sigma, \pi \in \mathcal{C}_p\}$. Following Theorem 1, we deduce that this basis \mathcal{B}_d of excluded patterns is made of the minimal permutations with $d = 2^p$ descents, that is to say the permutations with 2^p descents that contain no pattern with 2^p descents, except themselves. What is left to prove is that this basis is finite. We postpone this part of the proof

to Proposition 2, where an upper bound of the size of the patterns in \mathcal{B}_d is given. \square

1.4 Outline of the paper

In this paper, we focus on the basis \mathcal{B}_d of excluded patterns appearing in Theorem 2. More generally, we do not assume that d is a power of 2 and wish to characterize and enumerate the set \mathcal{B}_d of permutations that are the minimal ones in the sense of \prec for the property of having d descents.

The work that is presented hereafter is organized as follows. First, we give a local characterization of the permutations of \mathcal{B}_d . In Section 2, we prove that the permutations of \mathcal{B}_d are the permutations σ , whose ascents satisfy a simple and local property: there is an ascent in $\sigma \in S_n$ at position i if and only if $2 \leq i \leq n - 2$ and $\sigma_{i-1}\sigma_i\sigma_{i+1}\sigma_{i+2}$ forms an occurrence of either the pattern 2143 or the pattern 3142.

This characterization is used to try and count the permutations in \mathcal{B}_d . Despite our effort, we did not succeed in this direction, and focused on a simpler case that can be seen as a first step in the enumeration of \mathcal{B}_d . First, as explained at the beginning of Section 2, Proposition 2, the size of the permutations in \mathcal{B}_d is at least $d + 1$ and at most $2d$. Obviously there is only one permutation of \mathcal{B}_d of size $d + 1$, that is the reverse identity permutation $(d + 1)d(d - 1) \dots 321$. For any other size, there is no immediate result. Using a representation of permutations of \mathcal{B}_d as posets (partially ordered sets), we could enumerate the permutations in \mathcal{B}_d having size $2d$. In Section 3, we give the details of the proof that the permutations of \mathcal{B}_d having size $2d$ (*i.e.* maximal size) are enumerated by the Catalan numbers.

Section 4 lists some open problems in the study of the sets \mathcal{B}_d 's.

From here on, by minimal permutation with d descents, we mean a permutation that is minimal in the sense of the pattern-involvement relation \prec for the property of having d descents.

Example 3, which is illustrated on Figure 3, should clarify the notion of minimal permutation with d descents.

Example 3. *Permutation $\sigma = 8613241195107$ has 6 descents but is not minimal with 6 descents. Indeed, the elements 1 and 4 (that are circled on Figure 3) can be removed from σ without changing the number of descents.*

Doing this, we obtain permutation $\pi = 642197385$ which is minimal with 6 descents: it is impossible to remove an element from it while preserving the number of descents equal to 6.

However, π is not of minimal size among the permutation with 6 descents: π has size 9 whereas permutation 7654321 has 6 descents but size 7.

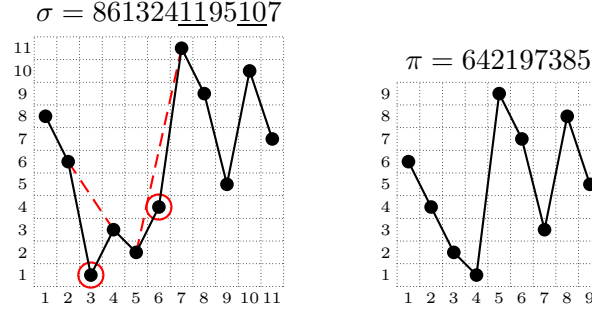


Figure 3: Permutations σ and π of Example 3.

2 A characterization for minimal permutations with d descents

The aim of this section is to provide a more practical characterization of minimal permutations with d descents, by finding necessary and sufficient conditions on permutations for being minimal with d descents. First, we provide a necessary condition on the size of those permutations with Proposition 2.

Proposition 2. *Let σ be a minimal permutation with d descents. Then every ascent of σ is immediately preceded and immediately followed by a descent, and the size n of σ satisfies $d + 1 \leq n \leq 2d$.*

Proof. Consider a permutation $\sigma \in \mathcal{B}_d$, and denote by n the size of σ . By minimality in the sense of \prec , σ has exactly d descents. To create a permutation with d descents, you need at least $d + 1$ elements, and with $d + 1$ elements, the only permutation with d descents you can create is $(d + 1)d(d - 1) \dots 21$, which is minimal. Therefore, $n \geq d + 1$.

It is also easily seen that σ neither starts nor ends with an ascent, otherwise the permutation obtained by removing the first or the last element of σ would have the same number d of descents, contradicting that σ is minimal with d descents. In the same way, σ cannot have two consecutive ascents $\sigma_{i-1}\sigma_i$ and $\sigma_i\sigma_{i+1}$, otherwise we would get the same contradiction removing σ_i , since this removal does not change the number of descents.

This proves that a minimal permutation with d descents is composed of non-empty sequences of descents, separated by isolated ascents. A longest possible permutation with d descents so obtained has d isolated descents, separated by $d - 1$ isolated ascents, and consequently has $2d$ elements. We then get that the size of σ is at most $2d$: $n \leq 2d$. \square

The decomposition of a minimal permutation with d descents into non-empty sequences of descents separated by isolated ascents that is described

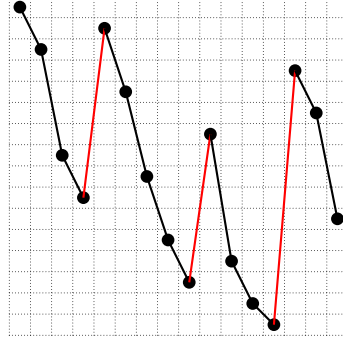


Figure 4: Decomposition of a minimal permutation with d descents into non-empty sequences of descents separated by isolated ascents

in the proof of Proposition 2 is illustrated in Figure 4. This decomposition can be carried further to give a necessary and sufficient condition on permutations for being minimal with d descents. This characterization is described in Theorem 3.

Theorem 3. *A permutation σ is minimal with d descents if and only if it has exactly d descents and its ascents $\sigma_i\sigma_{i+1}$ are such that $2 \leq i \leq n-2$ and $\sigma_{i-1}\sigma_i\sigma_{i+1}\sigma_{i+2}$ forms an occurrence of either the pattern 2143 or the pattern 3142.*

Proof. Let σ be a minimal permutation with d descents. In the decomposition of σ into non-empty sequences of descents separated by isolated ascents – illustrated in Figure 4 – it appears clearly that an ascent $\sigma_i\sigma_{i+1}$ is necessarily such that $2 \leq i \leq n-2$, with $\sigma_{i-1}\sigma_i$ and $\sigma_{i+1}\sigma_{i+2}$ being descents.

Now, consider an ascent $\sigma_i\sigma_{i+1}$. The previous remarks lead to $\sigma_{i-1} > \sigma_i$, $\sigma_i < \sigma_{i+1}$ and $\sigma_{i+1} > \sigma_{i+2}$.

Let us assume that $\sigma_{i-1} > \sigma_{i+1}$. Then the permutation obtained from σ by the removal of σ_i has as many descents as σ (and one ascent less), contradicting the minimality of σ . Consequently, $\sigma_{i-1} < \sigma_{i+1}$. Similarly, we get that $\sigma_i < \sigma_{i+2}$ (see Figure 5).

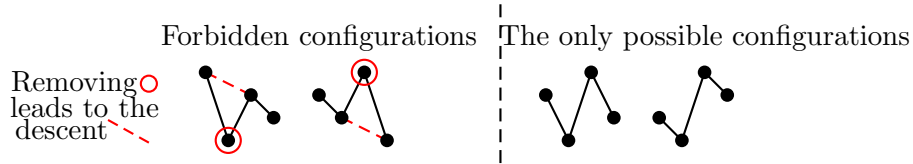


Figure 5: The elements $\sigma_{i-1}\sigma_i\sigma_{i+1}\sigma_{i+2}$ around an ascent $\sigma_i\sigma_{i+1}$ in a permutation σ which is minimal with d descents

At this point, we have the five following inequalities: $\sigma_{i-1} > \sigma_i$, $\sigma_i < \sigma_{i+1}$, $\sigma_{i+1} > \sigma_{i+2}$, $\sigma_{i-1} < \sigma_{i+1}$ and $\sigma_i < \sigma_{i+2}$. Thanks to them it is possible

to check that the sequence $\sigma_{i-1}\sigma_i\sigma_{i+1}\sigma_{i+2}$ is an occurrence of either the pattern 2143 or the pattern 3142.

Conversely, consider a permutation σ with d descents whose ascents $\sigma_i\sigma_{i+1}$ are such that $2 \leq i \leq n-2$ and $\sigma_{i-1}\sigma_i\sigma_{i+1}\sigma_{i+2}$ forms an occurrence of either the pattern 2143 or the pattern 3142. This implies that σ has the shape of non-empty sequences of descents separated by isolated ascents. And it is a simple matter to prove that the removal of any element of σ makes the number of descents decrease by one – there are three cases to consider: the removed element may be either the first element of an ascent, or the second element of an ascent, or it may be between two descents. This proves that σ is a minimal permutation with d descents. \square

We thought this characterization could help us to enumerate the minimal permutations with d descents. Although we did not reach this goal, we obtain partial results studying minimal permutation with d descents and of a given size n . For $n = d + 1$, we already proved that there is only one such permutation. For $n = 2d$, the next section describes the enumeration we obtained. We will use a *partially ordered set* (or *poset*) representation of permutations, that comes directly from the characterization of minimal permutations with d descents in Theorem 3.

Representation of minimal permutations with d descents with posets

Consider a set of all the permutations of a given size n , that are minimal with d descents, and having their descents and ascents in the same positions. In all these permutations, the elements are locally ordered in the same way, even around the ascents, because of Theorem 3. We can give a representation of this whole set of permutations by a *partially ordered set* (or *poset*) indicating the necessary conditions on the relative order of the elements between them. For a descent, we just have a link from the first and greatest element to the second and smallest one. For any ascent $\sigma_i\sigma_{i+1}$, the elements $\sigma_{i-1}\sigma_i\sigma_{i+1}\sigma_{i+2}$ form a diamond-shaped structure with σ_{i+1} on the top, σ_i on the bottom, σ_{i-1} on the left and σ_{i+2} on the right. See Figure 6 for an example. By Theorem 3, any labelling of the elements of the poset respecting its ordering constraints is a minimal permutation with d descents.

3 Enumeration of minimal permutations with d descents and of size $2d$

The minimal permutations with d descents that have size $2d$ are, because of minimality, of a very particular shape. Indeed, they cannot have two consecutive ascents as usual, but neither can they have two consecutive descents, otherwise it would be impossible to reach size $2d$. Consequently, they all result from an alternation of isolated descents and isolated ascents,



Figure 6: A poset representing a set of minimal permutations with 16 descents and 4 ascents (consequently of size 21) containing, among others the permutation 20 18 15 14 19 17 10 8 13 12 21 16 11 9 7 5 3 2 6 4 1 whose grid representation is also given

of course starting and ending with a descent. An example is given in Figure 7(a).

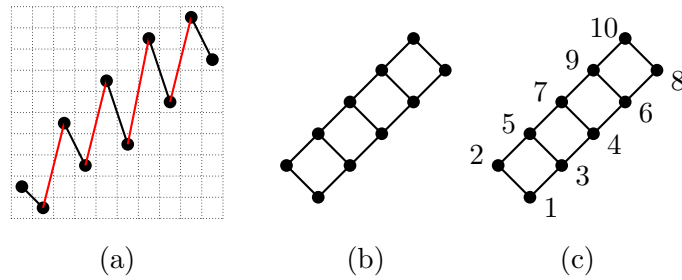


Figure 7: (a) A minimal permutation $\sigma = 2\ 1\ 5\ 3\ 7\ 4\ 9\ 6\ 10\ 8$ with $d = 5$ descents and of size $2d = 10$, (b) the poset representing the set of all minimal permutations with $d = 5$ descents and of size $2d = 10$ and (c) the authorized labelling of the subsequent poset associated with σ

A consequence is that all minimal permutations of size $2d$ with d descents have their descents and ascents in the same position, so that a unique poset represents the set of all minimal permutations with d descents having size $2d$. This poset has the shape of a ladder with d steps: it is a sequence of $d - 1$ diamonds, two consecutive diamonds being linked by an edge. These diamonds correspond to the ascents in the permutations, that are separated by one descent only in this case. See Figure 7(b) for an example.

The paragraph on poset representation at the end of Section 2 justifies Proposition 3:

Proposition 3. *The minimal permutations with d descents and of size $2d$ correspond exactly to the labellings of the ladder poset with d steps with the integers $\{1, 2, \dots, 2d\}$ that respect its ordering constraints.*

An example of this correspondance is given in Figure 7(c).
The main result of this section is:

Theorem 4. *The minimal permutations with d descents and of size $2d$ are enumerated by the Catalan numbers $C_d = \frac{1}{d+1} \binom{2d}{d}$.*

Proof of Theorem 4 by enumeration: By using the ECO method, detailed in [3], we build all the authorized labellings of the ladder poset with d steps from all the authorized labellings of the ladder poset with $d - 1$ steps by a process of local expansion, and without creating twice the same labelling.

In the ECO method, the combinatorial objects (labellings of the ladder poset with d steps in our case) receive *labels*. The label of an object is the number of its *children*, that is to say the number of objects that are obtained from it in the local expansion process. Those children can again receive a label by the same method.

With the ECO labelling of the combinatorial objects, we derive a *succession rule* or *rewriting rule* that describes the production (in terms of labels) of the possible labels of these objects, together with a starting point.

A possible way of proving that authorized labellings of the ladder poset with d steps are enumerated by the Catalan numbers is to show an ECO construction for this class whose associated succession rule is known to correspond to the Catalan numbers.

The ECO labels that are given to authorized labellings of the ladder posets with d steps for this purpose are $(2d - \sigma_{2d} + 1)$, σ_{2d} being the label of the rightmost element of the poset. Notice also that $2d$ is the label of the uppermost element of the poset, and that this element is also the second rightmost one.

Consider an authorized labelling σ of the ladder poset with d steps that has ECO label (k) . Its children are the labellings of the ladder poset with $d + 1$ steps obtained by adding a new step on the right, this new step of the ladder being labelled with $2d + 2$ for the top element, and i for the rightmost one, for $2d + 2 - k \leq i \leq 2d + 1$. The elements j in σ with $j \geq i$ are turned into $j + 1$ to maintain both the relative order of the elements of σ and the property that these new labellings use all the integers of $[1..2d + 2]$ exactly once.

Since $k = 2d - \sigma_{2d} + 1$, it is easy to check that all the labellings obtained in this way are authorized, and that all of them are obtained. The ECO labels of the k children (of size $d + 1$) of an authorized labelling of size d with ECO label (k) are, by the above formula, $(2(d + 1) - i + 1)$ with $2d + 2 - k \leq i \leq 2d + 1$, that is to say the children of a labelling with ECO label (k) have labels $(2), (3), \dots, (k), (k + 1)$.

The starting point for this ECO construction is the ladder poset with one step provided with its only authorized labelling 21, and whose ECO

label is (2).

To sum up, the succession rule obtained for this ECO construction of authorized labelling of the ladder posets is

$$\left\{ \begin{array}{l} (2) \\ (k) \rightsquigarrow (2)(3) \cdots (k)(k+1) \end{array} \right.$$

and this succession rule is known to correspond to combinatorial classes enumerated by Catalan numbers (Dyck paths among others).

Proof of Theorem 4 by bijection: It is well known that Dyck paths of length $2d$ are enumerated by the Catalan numbers $C_d = \frac{1}{d+1} \binom{2d}{d}$.

We provide a bijection between Dyck paths of length $2d$ and authorized labellings of the ladder poset with d steps with the integers $\{1, 2, \dots, 2d\}$. The bijection is simple. Starting from a Dyck path \mathcal{D} of length $2d$, we number its steps with the integers from 1 to $2d$, from left to right. Then, we label the lower line of the ladder with the numbers of the steps of \mathcal{D} going up and its upper line with the numbers of the steps of \mathcal{D} going down. An example is shown in Figure 8.

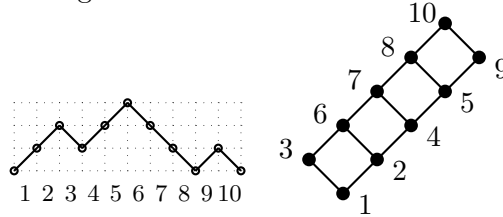


Figure 8: An example of the bijection between minimal permutations with d descents of size $2d$ (seen as authorized labellings of the ladder poset with d steps) and Dyck paths with $2d$ steps

The application we described is actually a bijection between Dyck paths and the authorized labellings of the ladder posets, corresponding to the permutations we are interested in. The reason is simple. It is sufficient to notice that a labelling of the ladder poset with d steps is authorized if and only if the two lines of the ladder are increasing sequences and any i -th element x on the upper line has at least i smaller elements on the lower line (the element y on the lower line that is linked to x by a step on the ladder, and all the elements below y). See Figure 9 for a better understanding of this statement. In the same way, a path with d steps going up and d steps going down is a Dyck path if and only if any i -th step going down has at least i steps going up before it.

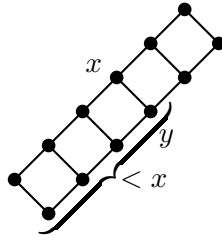


Figure 9: A condition for a labelling of the ladder poset to be authorized

4 Conclusion and open problems

The goal pursued in this paper is the analysis (characterization, enumeration, ...) of the permutations that are minimal for the property of having d descents, minimal being intended in the sense of the pattern-involvement relation. For $d = 2^p$, those permutations arise from the whole genome duplication - random loss model, defined in computational biology, where they appear as the excluded patterns defining the pattern-avoiding classes of permutations obtained in at most p steps in this model.

We first provided a local characterization of the minimal permutations with d descents, focusing only on the elements of the permutation surrounding its ascents. This characterization is easy to check: indeed, it provides a linear-time procedure for deciding whether a permutation is minimal with d descents or not.

The second step of our study was more about enumerating these permutations. We proved that a minimal permutation with d descents has size at least $d + 1$ and at most $2d$. We could not find the enumeration of all minimal permutations with d descents, but we were able to enumerate such permutations of size $d + 1$ and $2d$. More precisely, there is only one of size $d + 1$ (which is the reversed identity) and those of size $2d$ are enumerated by the Catalan numbers.

The enumeration of the minimal permutations with d descents and of size $n \in [(d + 2)..(2d - 1)]$ remains an open question. Moreover we computed the first few terms of the enumerating sequence for $n = d + 2$ and it appears in the Online Encyclopedia of Integer Sequences [11] (twice the sequence [A002662]). Therefore it could be interesting to find a bijective enumeration of this set of permutations.

References

- [1] M.H. Albert, R.E.L. Aldred, M.D. Atkinson, H.P. Van Ditmarsch, C.C. Handley, D.A. Hotlon, and D.J. McCaughan. Compositions of pattern restricted sets of permutations. Technical report, 2004.

- [2] M. D. Atkinson, M. M. Murphy, and N. Ruskuc. Partially well-ordered closed sets of permutations. 2001.
- [3] E. Barcucci, A. Del Lungo, E. Pergola, and R. Pinzani. ECO: A methodology for the enumeration of combinatorial objects. *J. Difference Equ. Appl.*, 5:435–490, 1999.
- [4] S. Bérard, A. Bergeron, C. Chauve, and C. Paul. Perfect sorting by reversals is not always difficult. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 4(1), 2007.
- [5] A. Bernini, L. Ferrari, and R. Pinzani. Enumerating permutations avoiding three Babson-Steingrímsson patterns. *Annals of Combinatorics*, 9:137–162, 2005.
- [6] M. Bousquet-Mélou. Four classes of pattern-avoiding permutations under one roof: generating trees with two labels. *Electronic Journal of Combinatorics*, 9, 2003.
- [7] M. Bouvel and D. Rossin. A variant of the tandem duplication - random loss model of genome rearrangement. arXiv:0801.2524v1 [math.CO].
- [8] K. Chaudhuri, K. Chen, R. Mihaescu, and S. Rao. On the tandem duplication-random loss model of genome rearrangement. SODA, pages 564 – 570, 2006.
- [9] M.C. Chen and R.C.T. Lee. Sorting by transpositions based on the first increasing substring concept. In *BIBE '04: Proceedings of the 4th IEEE Symposium on Bioinformatics and Bioengineering*, page 553, Washington, DC, USA, 2004. IEEE Computer Society.
- [10] A. Labarre. New bounds and tractable instances for the transposition distance. *IEEE/ACM Trans. Comput. Biology Bioinform.*, 3(4):380–394, 2006.
- [11] N. J. A. Sloane. The On-Line Encyclopedia of Integer Sequences, 2007. published electronically at www.research.att.com/~njas/sequences/.
- [12] V. Vatter. Enumeration schemes for restricted permutations. Permutation patterns, 2005.